

TV 광고 식별을 위한 Constant-Q 변환 기반의 오디오 핑거프린팅 방식

Audio Fingerprinting Based on Constant Q Transform for TV Commercial Advertisement Identification

류상현, 김형국[†]

(Sang Hyeon Ryu and Hyoung-Gook Kim[†])

광운대학교 전자공학과

(접수일자: 2013년 12월 30일; 수정일자: 2014년 2월 10일; 채택일자: 2014년 3월 18일)

초 록: 오디오 핑거프린팅 기술은 잡음과 에코 등으로 인한 왜곡에도 성공적으로 음원을 식별해야 한다. 이러한 오디오 핑거프린팅 기술을 TV 광고 식별에 적용하고자 한다. 본 논문은 TV 광고 식별을 위한 강인한 오디오 핑거프린팅 방식을 제안한다. 제안된 방법에서 사용되는 Constant Q 변환 기반에서 추출된 현저한 오디오 피크 쌍 핑거프린트는 실제 다양한 잡음환경에서 오디오 핑거프린팅 시스템의 정확도를 향상시키고, 낮은 복잡도를 가진다. 실험 결과는 제안된 방식이 기존의 오디오 핑거프린팅 방식에 비해 다양한 잡음환경에서도 안정적이며 신뢰할 수 있는 검색 정확도를 제공함을 보여준다.

핵심용어: 오디오 핑거프린팅, Constant Q 변환, 콘텐츠 식별, 강인한 매칭

ABSTRACT: In spite of distortion caused by noise and echo, the audio fingerprinting technique must identify successfully an audio source. This audio fingerprinting technique is applying for TV commercial advertisement identification. In this paper, we propose a robust audio fingerprinting method for TV commercial advertisement identification. In the proposed method, a prominent audio peak pair fingerprint based on constant Q transform improves the accuracy of the audio fingerprinting system in real noisy environments. Experimental results confirm that the proposed method is quite robust than previous audio fingerprinting method in different noise conditions and achieves promising accurate results.

Keywords: Audio fingerprinting, Modulation constant Q transform, Content identification, Robust matching

PACS numbers: 43.60.Hj, 43.60.Lq, 43.75.Cd

1. 서 론

오디오 핑거프린팅 기술은 배경잡음, 에코 등에 의한 왜곡에도 불구하고 몇 초 동안의 짧고, 표기되지 않은 오디오 클립을 성공적으로 식별하는 기능을 제공하는 기술로서 검색 정확도를 향상시키기 위해서 다양한 방법^[1]들이 연구 및 개발되어 오고 있다.

이러한 오디오 핑거프린팅 기술은 음악식별을 기

반으로 음악서비스에 활용되어 오다가 최근 들어 모바일 기기를 통한 TV 광고식별 분야에 적용되고 있다. 특별히, Wang^[2]의 오디오 핑거프린팅 시스템과 Haitma and Kalker^[3]가 제안한 해시 알고리즘은 대표적인 오디오 핑거프린팅 기술로서 높은 정확도를 제공하고 있기 때문에, 현재 이를 응용 및 개선한 다양한 알고리즘이 개발되어 모바일 기기에 구현되어 오고 있다.

Wang의 방법에서는 각 오디오 트랙으로부터 에너지가 집중된 주파수의 두드러진 피크를 찾기 위해 Short Time Fourier Transform(STFT)를 사용한다. 이 피

[†]Corresponding author: Hyoung-Gook Kim (hkim@kw.ac.kr)
Department of Wireless Communication Engineering, Kwang-Woon University, 20 Gwangun-Ro, Nowon-Gu, Seoul 139-701, Republic of Korea
(Tel: 82-2-940-5574, Fax: 82-2-913-5006)

크들은 주파수와 시간에 의해서 파라미터화된 타겟 영역 내에서 피크들 간의 쌍으로 형성된다. 형성된 피크 쌍은 쿼리를 식별하기 위해, 양자화 되어 landmark로 변환하고 해시로 사용한다. 쿼리에서 추출된 해시의 시간차 정보와 데이터베이스의 핑거프린트의 시간차 정보를 비교하여 가장 유사한 매칭결과를 획득한다. Wang 알고리즘의 아이디어를 기반으로, Pan^[4]은 오디오 핑거프린트를 생성하기 위한 지역 에너지 centroid 방식, Jiang^[5]은 실시간 피크 검색 방식을 제안하였다.

Haitsma and Kalker의 방법에서는 오디오 트랙으로부터 오버랩된 각 프레임의 스펙트럼을 300~2000 Hz 사이의 33개 서브밴드로 나누고, 인접한 서브밴드간의 에너지 차이관계에 의해서 총 32 bit의 핑거프린트를 추출한다. 추출된 핑거프린트의 고속 검색을 위해서는 32 bit 값을 이용한 lookup table을 이용하였다. 이 방식의 결점은 TV 광고 음원을 식별하기 위한 정보의 양이 상대적으로 많고 낮은 신호대 잡음비(SNR)에서 낮은 성능을 보인다.^[4] Haitsma and Kalker의 알고리즘을 기반으로 Park^[6]은 실제 잡음 환경에서 견고한 핑거프린팅 방법을 위해 주파수 시간 필터링, Son^[7]은 Haitsma and Kalker의 해싱 알고리즘의 각 서브 핑거프린트에서 두드러진 피치 추정에 의해 생성된 마스크 방식, 그리고 Anguera^[8]은 Wang과 Haitsma and Kalker의 아이디어를 기반으로, 견고한 오디오 핑거프린팅을 위해 마스크 된 오디오의 스펙트럼 키폰트 방식을 제안하였다.

TV 광고 식별을 위한 오디오 핑거프린팅 시스템의 정확도를 향상시키기 위해서는 무엇보다도 잡음 환경과 왜곡에 강인한 오디오의 특성을 반영한 오디오 핑거프린트를 추출해야 한다. 본 논문에서는, Wang의 방법의 아이디어를 기반으로, 실제 잡음 환경에서 오디오 기반 TV 광고 식별 시스템의 성능을 향상시키기 위해 Constant Q Transform(CQT)기반의 현저한 피크 검출을 통한 오디오 핑거프린팅 방법을 제안한다.

본 논문은 다음과 같이 구성되었다. 2장에서 제안된 방법을 설명한다. 3장에서 실험결과를 제시한다. 마지막으로, 4장에서 결론을 서술한다.

II. Constant Q 변환 기반의 오디오 핑거프린팅

제안된 CQT 기반의 현저한 오디오 피크 쌍 핑거프린팅 방식은 Fig. 1과 같이, 다섯 단계로 구성된다. 먼저, 전처리 과정으로, 입력된 스테레오 오디오 신호는 모노로 변환되고(스테레오 좌우 오디오 신호의 각 샘플값별 평균값을 모노신호로 변환), 16kHz로 다운샘플링 된다. 변환된 오디오 신호는 Hamming Window 함수(각각 512개 샘플 오버랩)에 의해 오버랩된 프레임으로 나뉜다. 스펙트럼 피크를 찾기 위해, CQT가 각 프레임(1024개 샘플)마다 적용된다. 세부적인 단계별 과정은 다음과 같다.

2.1 CQT 스펙트럼 획득과 Log 스케일 평균 차감

입력된 오디오 신호 $x(n)$ 은 Hamming window된 오버랩 프레임으로 나뉘고, CQT를 사용하여 오디오 신호의 스펙트럼을 분석한다.

$$X_{CQT}(k,l) = \frac{1}{N_k} \sum_{n=0}^{N_k-1} w(n,k)x(n+lM)e^{-\frac{j2\pi Qn}{N}}, \quad (1)$$

여기서, k 는 주파수 빈 인덱스, l 은 시간 프레임 인덱스, w 는 N_k 크기의 window 분석[$w(n,k)$ 의 크기 N_k 는 주파수 빈 별로 변화된다], M 은 프레임 이동 단계이

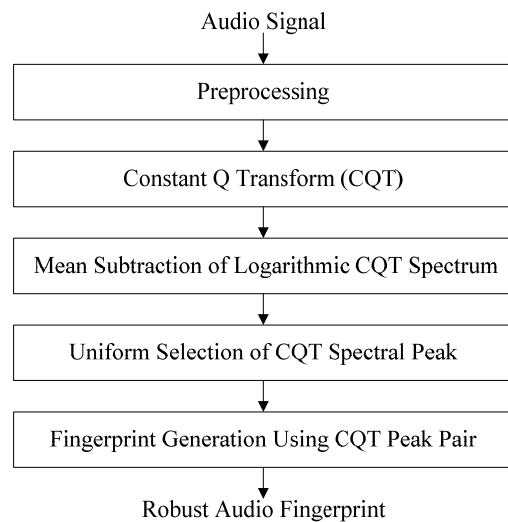


Fig. 1. Block diagram of the robust audio fingerprint extraction.

며, $Q(45 < Q < 55)$ 는 중심 주파수와 주파수 대역의 비율로 계산되는 상수로, 각 주파수 대역의 크기를 조절하여 해상도를 조절하는 품질 상수이다.

CQT 기반의 스펙트럼 분석은 낮은 주파수 영역에서는 좁은 대역폭을 가지고, 높은 주파수 영역에서는 넓은 대역폭을 가지는, 인간 귀의 주파수 특성과 비슷한 특성을 가지고 있다. STFT 기반의 방식과 같이 낮은 주파수 영역의 해상도를 높은 주파수 영역에 적용하는 것은 시스템적인 낭비이다. 그래서 CQT에서는 주파수마다 해상도를 달리하여 낮은 주파수대역에서는 STFT에 비해 좁은 영역을 분석함으로써 주파수특성에 대한 이점을 살려 더 많은 피크를 보존할 수 있고, 높은 주파수 영역에서는 STFT에 비해 넓은 영역을 한번에 분석함으로써 더 효과적으로 주파수 피크를 식별 할 수 있다.

이전 프레임에서 검색된 최소값과 CQT 스펙트럼의 비교를 사용하여, 모든 프레임에서 구역 에너지의 최소값 $X_{\min}(k, l)$ 을 찾는다. 찾아진 최소값으로부터, 각 주파수 빈에서 $X_{\min}(k, l)$ 의 최대값을 얻는다.

$$X_{\max}(k, l) = \max [X_{CQT}(k, l), X_{\min}(k-1, l)] \quad (2)$$

$$, 0 \leq k < K .$$

CQT 스펙트럼과 $X_{\max}(k, l)$ 의 비교를 통해 각각의 CQT 계수에 Log 스케일을 다음과 같이 수행한다.

$$\begin{aligned} \text{If } [X_{CQT}(k, l) > X_{\max}(k, l)] \\ X_{\log}(k, l) = \log_{10} [X_{CQT}(k, l)] \\ \text{else} \\ X_{\log}(k, l) = \log_{10} [X_{\max}(k, l)] \end{aligned} \quad (3)$$

Log 스케일의 CQT 스펙트럼 $X_{\log}(k, l)$ 로부터, CQT 스펙트럼의 평균 X_{mean} 이 추정되고, Log 스케일 CQT 스펙트럼의 하한 및 상한의 ripple을 최소화 하기 위해 모든 프레임에서 차감된다.

$$X_{norm}(k, l) = X_{\log}(k, l) - X_{mean} \quad (4)$$

$$X_{mean} = \frac{1}{K \cdot L} \sum_{l=0}^{L-1} \sum_{k=0}^{K-1} X_{\log}(k, l) \quad (5)$$

2.2 CQT 스펙트럼 피크 검출

정규화된 Log 스케일의 CQT 스펙트럼 $X_{norm}(k, l)$ 에서 어느 한 점을 중심으로 주위 영역보다 높은 에너지를 포함한다면 이는 피크의 후보군이라 할 수 있다. 피크 후보군은 정규화된 Log 스케일의 CQT 스펙트럼을 통해 분석된 시간-주파수 영역에서 잡음 혹은 왜곡에도 강인하다. 이러한 피크후보군을 검출하기 위해서 적응적인 피크 검출 문턱값을 사용하는 순방향과 역방향 필터링 방식이 적용되며, 그 과정은 다음과 같다.

2.2.1 초기 순방향 문턱값 계산

정규화된 Log 스케일의 CQT 스펙트럼과 J 프레임 범위에서 이전 프레임의 최대값과 프레임 단위 비교를 통해서 각 주파수 빈의 최대값을 산출한다.

$$F_{Max}(k, 0) = \max [X_{norm}(k, l), F_{Max}(k, l-1)] \quad (6)$$

$$, 0 \leq l < J .$$

기울기가 양에서 음으로 변화가 발생하는 기본적인 가정을 사용하여, $F_{Max}(k, l)$ 에서 극대값 $F_S(k, l)$ 을 선택한다. 극대값이 선택되지 않은 부분은, 인접한 2개의 극대값 간에 값이 선형적으로 보간되어, 첫 번째 프레임을 위한 문턱값 $Th_f(k, 0)$ 으로 사용된다.

$$\begin{aligned} \text{If } [F_S(k_{pre}, l) > F_S(k, l)] \\ Th_f(k, 0) = F_S(k_{pre}, 0) - \frac{F_S(k_{pre}, 0) - F_S(k, 0)}{k - k_{pre}} k' \\ \text{else} \\ Th_f(k, 0) = F_S(k_{pre}, 0) + \frac{F_S(k, 0) - F_S(k_{pre}, 0)}{k - k_{pre}} k' \end{aligned} \quad (7)$$

여기서 $k'(k_{pre} < k' < k)$ 는 k_{pre} 와 k 사이의 주파수 인덱스이고, k_{pre} 는 극대값에서 찾아진 이전 주파수 인덱스이다.

2.2.2 순방향 최대 피크(Forward High Peak, FHP)

선정과 적응적인 피크 검출 문턱값 갱신

순방향 문턱값 $Th_f(k, l)$ 보다 큰 $X_{norm}(k, l)$ 의 모든 피크가 순방향 최대 피크라는 명칭의 (k, l) 의 형태로 저장된다.

$$X_{norm}(k,l) = \begin{cases} FHP, & \text{if } X_{norm}(k,l) > Th_f(k,l) \\ non\ FHP, & \text{otherwise} \end{cases} \quad (8)$$

만약 $X_{norm}(k,l)$ 중에서 FHP 가 선택되면, FHP 는 $P_f(k,l)$ 로 표현되고, 피크 검출 문턱값이 갱신된다.

$$\begin{aligned} & \text{If } [P_f(k_{pre},l) > P_f(k,l)] \\ & Th_{Up}(k,l) = P_f(k_{pre},l) - \frac{P_f(k_{pre},l) - P_f(k,l)}{k - k_{pre}} k' \\ & \text{else} \\ & Th_{Up}(k,l) = P_f(k_{pre},l) + \frac{P_f(k,l) - P_f(k_{pre},l)}{k - k_{pre}} k'. \end{aligned} \quad (9)$$

다음 프레임을 위한 새로운 피크 검출 문턱값이 감쇠 계수 $d(k,l)$ 와 갱신된 피크 검출 문턱값 $Th_{Up}(k,l)$, 그리고 감쇠 이전 문턱값을 비교를 통해 적응적으로 계산된다.

$$\begin{aligned} & Th_f(k,l) \\ & = \max [d(k,l-1) \cdot Th_f(k,l-1), Th_{Up}(k,l)], \end{aligned} \quad (10)$$

$$d(k,l-1) = \exp \left(\lambda \cdot \left| \frac{Th_f(k,l-1) - m}{\sigma} \right| \right), \quad -1 \leq \lambda \leq 1, \quad (11)$$

여기서 m 과 σ 은 각 문턱값의 주파수 밴드의 평균과 분산을 나타낸다. 감쇠 계수를 사용하는 문턱값은 원본 핑거프린트와 쿼리의 핑거프린트를 비교하기 위해 더 현저한 피크를 추출할 수 있다. $l = l+1$ 에 의해 모든 프레임이 처리될 때 까지 순방향 피크를 계속 검출한다.

2.2.3 역방향 최대 피크(Backward High Peak, BHP) 검출

앞서 검출한 FHP 를 검증하고 잡음에 강인한 피크를 추출하기 위해, 마지막 프레임에서부터 역방향으로 피크를 검출한다. BHP 검출에서는 $X_{norm}(k,l)$ 의 마지막 프레임의 값을 이용해서 BHP 검출의 초기 문턱값을 생성한다. 그리고 BHP 검출은 FHP 를 검증하는 단계이기 때문에, 앞서 검출하고 저장된 FHP 의 (k,l) 만을 고려한다.

$$P_f(k,l) = \begin{cases} peak, & \text{if } P_f(k,l) > Th_b(k,l) \\ non\ peak, & \text{otherwise} \end{cases}, \quad (12)$$

$$\begin{aligned} & Th_b(k,l) \\ & = \max [d(k,l-1) \cdot Th_b(k,l-1), Th_{Up}(k,l)], \end{aligned} \quad (13)$$

$$\begin{aligned} & \text{If } [P_b(k_{pre},l) > P_b(k,l)] \\ & Th_{Up}(k,l) = P_b(k_{pre},l) - \frac{P_b(k_{pre},l) - P_b(k,l)}{k - k_{pre}} k' \\ & \text{else} \\ & Th_{Up}(k,l) = P_b(k_{pre},l) + \frac{P_b(k,l) - P_b(k_{pre},l)}{k - k_{pre}} k', \end{aligned} \quad (14)$$

여기서 $P_b(k,l)$ 는 적응적인 피크 검출 문턱값 $Th_b(k,l)$ 을 사용하는 역방향 필터링 후에 검출된 피크이다.

2.3 CQT 피크 쌍을 사용하는 핑거프린트 해시

핑거프린트 해시들은 $P_b(k,l)$ 쌍의 시간-주파수 정보를 연관 지음으로써 생성된다. 각 피크 쌍 (landmark로 명칭)은 기준 점과 정의된 영역 내에서 근접하는 landmark를 결합하여 선택된다. $P_b(k_a, l_a)$ 는 기준 점이고 $P_b(k_p, l_p)$ 는 결합된 다른 landmark들이라 가정하여, 해시를 얻는다.

$$hash = (l_a, k_a, k_p - k_a, l_p - l_a) = (l_a, k_a, \Delta k, \Delta l). \quad (15)$$

모든 k (주파수 빈)과 l (프레임)은 고정된 높은 바운드와 정수이고, 각 landmark점은 쌍의 고정 숫자를 생성한다. 획득한 핑거프린트 해시에 의해 데이터베이스의 인덱스가 생성되고, track ID와 해시의 시간 오프셋이 빠른 처리속도를 위한 해시값에 저장된다.

2.4 핑거프린트 데이터베이스 식별

사용자가 알고자하는 TV 광고를 식별하기 위해 짧은 오디오 쿼리를 입력하여, 앞서 설명한 방식과 동일하게 쿼리의 핑거프린트를 생성한다. 생성된 쿼리의 핑거프린트에서 시간과 주파수의 오프셋 정보를 사용하여 쿼리의 해시를 획득한다.

이때, 중복된 해시를 제거하여 불필요한 비교를 막고 TV 광고 식별 정확도를 향상시킨다. 중복된 정보가 제거된 해시는 데이터베이스의 인덱스로 사용되어, 각 인덱스에 저장된 해시값을 불러온다.

해시값을 통해 track ID와 시간 오프셋을 얻는다. 먼저, 얻어진 track ID를 비교하여 가장 많은 수가 카

운트된 상위 20개의 음원 후보군 ID를 선정한다. 다음으로, 선정된 상위 20개의 음원 후보군에 대한 각각의 시간 오프셋을 비교하여, 동일한 시간 오프셋이 검색되는 수를 카운트한다. 카운트 수가 가장 많은 track ID가 쿼리와 가장 근접한 음원이라 판별하여, 최종적인 식별결과인 track ID를 출력한다.

III. 실험 결과

이 장에서는 TV 광고 식별을 위해 제안된 CQT 피크 쌍 핑거프린트 추출 알고리즘의 성능을 측정한다. 알고리즘의 성능을 네 가지 기존 방식들과 성능을 비교한다. 방식 1은 Wang^[2]이 제안한 STFT 기반의 피크 쌍 핑거프린트 추출방식, 방식 2는 피치 추출 기반의 서브 핑거프린트 마스킹 오디오 핑거프린트 방식,^[4] 방식 3은 마스킹된 오디오 스펙트럼 주요영역기반의 오디오 핑거프린트 추출 방식,^[8] 그리고 방식 4는 오디오 신호의 시간-크로마 표현을 적용한 오디오 핑거프린트 추출방식^[9]이다.

실험을 위해, 총 740h 길이의 4,000편의 TV 광고로 구성된 데이터베이스를 사용하였다. 각 광고는 10~15 s 길이로 구성된다. 모든 오디오 데이터는 모바일폰 혹은 스마트폰과 같은 휴대 기기를 고려하여 PCM 포맷, 모노, 16-bit, 16 kHz 샘플링 rate를 사용하였다. TV 광고 식별을 위한 오디오 쿼리 클립은 5s 길이로 구성되며, 모바일 전화기를 사용하여 2.1 채널 스피커와 연결된 TV로부터 5m 떨어진 거리에서 캡처되었다. 모바일 전화기에 내장된 핑거프린트 생성 모듈을 통해, 캡처된 오디오 클립으로부터 CQT 피크 쌍 핑거프린트가 추출되고, 서버 영역의 매칭 모듈로 쿼리가 전송되었다

임의로 생성한 3,000개의 쿼리는 서로 다른 수준의 다양한 형태를 가지는 잡음을 추가함으로써 생성되었다. 즉, 다섯 가지 다른 형태의 잡음(배틀 잡음, 움직이는 자동차 잡음, 백색 잡음, 길거리 잡음, 그리고 컴퓨터 팬 잡음)이 클립과 신호대잡음비 0 dB, 6 dB, 12 dB의 수준으로 첨가되었다. 오디오 쿼리 데이터는 각 세트마다 임의로 선택된 1,000개가 캡처되었고, 각 오디오 샘플은 임의의 세트 오프셋에서 30회 재생되었다.

Table 1. Comparative performance four schemes.

SNR	Averaged Recognition Rate (%)				
	CQT	MW ^[2]	MS ^[4]	MC ^[8]	MX ^[9]
clean	94.8	93.6	94.5	93.5	92.6
12 dB	93.7	89.3	92.8	86.2	75.4
6 dB	86.5	80.6	81.7	74.6	60.2
0 dB	74.1	67.5	67.9	58.5	53.4
Total	87.3	82.8	84.2	78.2	70.4

Table 1은 5s 길이의 쿼리를 사용한 네 가지 방식의 실험결과를 나타낸다. MW, MS, MC, MX는 각각 방식 1, 2, 3, 4를 나타내고, CQT는 제안된 방식이다. Table 1의 수치는 다섯 가지 다른 형태의 잡음환경에서의 실험에 대한 인식 결과의 평균을 나타내었다.

Table 1과 같이, 음악 식별에 대한 가장 좋은 인식 정확도는 94.8%이다. TV 광고는 오디오 신호에 묵음구간이 캡처되어 매칭에 사용되는 경우가 많기 때문에 더 높은 결과를 얻진 못했다. MW와 MS의 인식률이 매우 유사하지만, CQT 방식보다 낮음을 알 수 있다. MX는 가장 낮은 인식률을 보이고, SNR 0 dB에서 가장 나쁜 결과를 보였다. 제안된 방식은 평균 87% 이상의 만족스러운 성능을 보여주어, 기존의 다른 방식보다 뛰어난 성능을 보여줌을 알 수 있다. 또한, SNR 0 dB의 높은 잡음신호가 함께 입력되는 상황에서도 타 방식에 비해 높은 정확도인 74.1%의 결과를 보여주었다. 이 결과는, 실제 사용 환경에서 나타날 수 있는 다섯 가지의 잡음이 입력되는 상황을 모두 실험한 결과로, 실제 사용자가 음원을 식별하기 위해 사용하는 잡음환경에서도 좋은 성능을 보일 것으로 예상된다.

IV. 결 론

본 논문에서는 CQT라 불리는 오디오 신호의 수정된 스펙트럼 표현방법 기반의 강인한 오디오 피크 쌍 핑거프린트 추출을 제안하였고, 이를 통한 TV 광고 식별 성능을 측정하였다. 제안된 알고리즘은 CQT를 기반의 안정적인 강인한 피크 쌍 핑거프린트를 생성함으로써 Wang의 핑거프린트 알고리즘보다 잡음이 합성된 음원에서도 더 많은 음원 고유의 핑거프린트 정보와 양질의 핑거프린트 정보를 획득할

수 있기 때문에, 실제 잡음 환경에서 오디오 핑거프린팅 시스템의 정확도를 향상시켰다. 실험결과는 제안된 방법이 기존의 다른 방식과 비교하여 실제 사용자가 사용하는 환경에서도 더 좋은 결과를 보여주었으며, 많은 실제 휴대용 소비자 기기에 적합함을 보여주었다.

향후, 검색 알고리즘을 최적화 시키는데 초점을 둘 것이다. 그리고 더 강한 콘텐츠 식별을 위해 오디오와 비디오 핑거프린트를 결합한 방식을 연구할 예정이다. 제안된 방식은 앞으로 스마트 TV와 모바일 전화기에서 작동하는 콘텐츠 보안 어플리케이션에 적용 될 것이다.

감사의 글

본 연구는 미래창조과학부 및 정보통신산업진흥원의 대학IT연구센터육성 지원사업의 연구결과로 수행되었음(NIPA-2014-H0301-14-1019).

References

1. V. Chandrasekhar, M. Sharifi, and D. A. Ross, "Survey and evaluation of audio fingerprinting schemes for mobile query-by-example applications," in Proc. of 12th ISMIR, 801-806 (2011).
2. A. Wang, "An industrial strength audio search algorithm," in Proc. of 4th ISMIR, 7-13 (2003).
3. J. Haitsma, and T. Kalker, "A highly robust audio fingerprinting system," in Proc. of 3rd ISMIR, 107-115 (2002).
4. X. Pan, X. Yu, J. Deng, W. Yang, and H. Wang, "Audio fingerprinting based on local energy centroid," IET Int'l CCWMC, 351-354 (2011).
5. T. Jiang, R. Wu, J. Li, K. Xiang, and F. Dai, "A real-time peak discovering method for audio fingerprinting," in Proc. of 5th ICIMCS, 368-371 (2013).
6. M. Park, H. Kim, and S. Yang, "Frequency temporal filtering for a robust audio fingerprinting scheme in real-noise environment," J. ETRI, **28**, 509-512 (2006).
7. W. Son, H.-T. Cho, K. Yoon, and S.-P. Lee, "Sub-fingerprint masking for a robust audio fingerprinting system in a real noise environment for portable consumer devices," IEEE Trans. on CE, **56**, 156-169 (2010).
8. X. Anguera, A. Garzon, and T. Adamek, "MASK:

robust local feature for audio fingerprinting," IEEE ICME, 455-460 (2012).

9. M. Malekesmaeili, and R. K. Ward, "A novel local audio fingerprinting algorithm," IEEE 14th Int'l Workshop on MMSP, 136-140 (2012).

저자 약력

▶ 류 상 현(Sang-Hyeon Ryu)



2012년 2월: 광운대학교 전파공학과 학사
2012년 3월 ~ 현재: 광운대학교 전파공학과 석사과정

▶ 김 형 국(Hyung-Gook Kim)



2002년 7월: 독일 SIEMENS/Cortologic AG 책임연구원
2005년 3월: 독일 베를린 공과대학교 Assistant Professor
2007년 2월: 삼성종합기술원 수석연구원
2007년 3월 ~ 현재: 광운대학교 전파공학과 부교수